Molecular Heterogeneity Analysis and Prognostic Stratification of Driver-Gene mutated NSCLC Based on Concurrent Alteration Profiles

Jiamei Jin
Nankai University
Wenjin Pan
Nankai University
Yiming Liu
Chinese PLA General Hospital
Wenhan Cai
Chinese PLA General Hospital
Kai Zhao
Chinese PLA General Hospital
Zirui Zhu
Chinese PLA General Hospital
Herui Han
Chinese PLA General Hospital
Mingchuan Hu
Nankai University
Xiangming Qiu
Chinese PLA General Hospital
Xinran Tian
Nankai University
Jiaxin Wen
Chinese PLA General Hospital
Xinyang Li
Chinese PLA General Hospital
Yuan Liu
Chinese PLA General Hospital
Xizhe Zhang
Chinese PLA General Hospital
Zhiqiang Xue
xuezhiqiang301@126.com

Chinese PLA General Hospital

Article

Keywords:

Posted Date: December 3rd, 2025

DOI: https://doi.org/10.21203/rs.3.rs-8103198/v1

Additional Declarations: No competing interests reported.

Abstract

Comprehensive co-alteration profiling reveals distinct molecular subtypes and prognostic stratification in driver-gene-mutated non-small cell lung cancer (NSCLC). In this study, we systematically delineated the co-alteration profiles across eight different driver mutations and their subtypes using next-generation sequencing (NGS) data from 494 NSCLC patients. Unsupervised hierarchical cluster analysis based on high-frequency co-alterations was subsequently performed to identify molecular subtypes in surgical patients with classical EGFR mutations. The cluster analysis stratified patients into four distinct subgroups with significantly different relapse-free survival (RFS). Cluster 1, characterized by TP53 co-alterations, demonstrated the worst RFS in both the entire cohort and Stage Ⅱ subgroup, and emerged as an independent prognostic risk factor in multivariable analysis (Cluster 2 vs. 1: HR = 0.12, p < 0.001; Cluster 3 vs. 1: HR = 0.13, p < 0.001). The prognostic model effectively identified high-risk patients who might benefit from more intensive surveillance or adjuvant therapy. Our study established a co-alteration-based molecular subtyping as a powerful prognostic tool in EGFR-mutant NSCLC, providing critical insights for personalized treatment strategies.

Introduction

Oncogenic driver gene mutations represent pivotal molecular events in the pathogenesis and progression of non-small cell lung cancer (NSCLC). Over the past two decades, the rapid advancement of targeted therapies has fundamentally transformed the treatment landscape and clinical outcomes for NSCLC patients with driver gene mutations[1]. The initial success of epidermal growth factor receptor-tyrosine kinase inhibitors (EGFR-TKIs) paved the way for the subsequent discovery and clinical application of several targeted agents against a growing number of driver mutations, including ALK, KRAS, BRAF, and MET, thereby establishing driver mutation-based molecular subtype as the cornerstone of precision medicine in NSCLC management[2–4]. Concurrently, the rapid evolution of genetic testing technologies, particularly the widespread adoption of next-generation sequencing (NGS), has enabled comprehensive genomic profiling, identifying an increasing number of NSCLC patients harboring specific driver mutations and significantly accelerating the development of genotype-directed targeted therapies.

Nevertheless, significant heterogeneity in treatment response persists among patients harboring identical driver gene mutations. Some patients exhibit primary resistance, and some develop acquired resistance after a short-term response, while others maintain long-term drug sensitivity[5]. This clinical observation suggests that the primary driver mutations alone can not fully dictate the entire biological behavior of the tumor. Co-alterations, defined as additional genetic alterations coexisting with the primary driver mutations in tumor tissue, commonly involving genes such as TP53, STK11, PIK3CA, CDKN2A/B, MYC, and those related to DNA damage repair, cell cycle regulation, and RTK/RAS signaling pathways, are key contributors. These co-alterations can synergize with primary drivers to remodel intratumoral signaling networks, facilitating bypass pathway activation or amplifying downstream signaling, consequently influencing malignant phenotypes including cell proliferation, apoptosis evasion, and metabolic reprogramming. Ultimately, this molecular heterogeneity substantially contributes to diverse clinical outcomes[6]. Previous investigations have revealed that TP53, the most frequently co-altered tumor suppressor gene, is often associated with poorer prognosis[7–9], while STK11 co-alterations in KRAS-mutant lung cancer strongly correlates with immunotherapy resistance[10, 11]. Therefore, systematic characterization of co-alteration profiles in driver gene-positive NSCLC is essential for deepening our understanding of tumor heterogeneity, overcoming treatment resistance, and developing novel combination therapy strategies.

Although prior studies have analyzed the co-alteration profiles of specific driver mutations such as EGFR and KRAS, a systematic comparative analysis encompassing broad driver mutation spectra and their various subtypes is still lacking[12]. Furthermore, rtranslation of complex co-alteration information into clinically applicable prognostic tools requires further investigation. In Asian NSCLC patients, where EGFR mutations represent the most prevalent driver alterations, surgical resection remains the primary curative approach for early-stage disease. Yet, postoperative recurrence risk persists in substantial patient proportions. While the TNM stage provides fundamental prognostic stratification, it fails to fully capture recurrence risks inherent in tumor genomic characteristics. Therefore, leveraging NGS data from pre- or post-operative tumor samples to integrate co-alteration information, develop a molecular stratification model that surpasses TNM stage, accurately identify high-risk patients, and inform individualized adjuvant therapy decisions represents a highly valuable research direction.

Based on this background, our study aims to: 1) Utilize large-scale NGS data to comprehensively delineate co-alteration profiles across different driver mutations and their subtypes, revealing their patterns and disparities; 2) Conduct unsupervised cluster analysis based on co-alteration profiles to explore molecular subtypes in surgical patients with classical EGFR mutations 3) Evaluate the postoperative relapse-free survival (RFS) and overall survival (OS) of different co-alteration subtypes to validate their prognostic value independent of clinicopathological factors. We aim to provide novel insights into the tumor heterogeneity and evolutionary mechanisms across different mutation-driven NSCLC subtypes, and to establish foundations for applying co-alteration profiles in precise prognostic assessment and clinical decision-making. The design of our study was shown in Fig. 1.

Results

Patients’ characteristics

A total of 494 NSCLC patients harboring at least one predefined driver gene mutation were ultimately included in the analysis. The baseline characteristics of the study cohort are summarized in Table 1. The distribution of driver gene mutations revealed EGFR to be the most prevalent alteration (349 cases, 70.65%), followed by KRAS (52 cases, 10.53%), HER2 (33 cases, 6.68%), ALK (24 cases, 4.86%), MET (20 cases, 4.05%), RET (17 cases, 3.44%), BRAF (16 cases, 3.24%), and ROS1 (8 cases, 1.62%).

Table 1
Baseline characteristics of 494 enrolled patients

	Total(n = 494)
Age(years)	56.7 ± 10.1
Gender	
Female	283(57.3%)
Male	210(42.5%)
Sample source	
Surgery	451(91.3%)
Needle biopsy	26(5.3%)
Transbronchil biopsy	12(2.4%)
Pathological type	
Adenocarcinoma	480(97.1%)
Squamous carcinoma	3(0.6%)
Other NSCLC	5(1.0%)
Stage	
Stage Ⅰ	383(77.5%)
Stage Ⅱ	34(6.8%)
Stage Ⅲ	29(5.8%)
Stage Ⅳ	41(8.2%)
Driver mutation	
EGFR	349(70.65%)
KRAS	52(10.53%)
BRAF	16(3.24%)
ALK	24(4.86%)
ROS1	8(1.62%)
RET	17(3.44%)
MET	20(4.05%)
HER2	33(6.68%)

Analysis of RFS across different co-alteration clusters

The analysis of RFS analysis across the four co-alteration-based clusters revealed significant prognostic differences (Fig. 5). In the entire cohort, Cluster 1 was associated with the shortest RFS and poorest prognosis (Fig. 5A). Compared to Cluster 1, both Cluster 2 and Cluster 3 showed a significantly reduced risk of recurrence (Fig. 5B and Table 2; adjusted HR = 0.12 and 0.13, respectively; both $p < 0.001$). Cluster 4 alsodemonstrated a non-significant trend toward lower risk relative to Cluster 1 (Fig. 5B and Table 2; adjusted HR = 0.37, $p = 0.058$).

This significant difference persisted in the Stage Ⅰ patient subgroup (Fig. 5C, Log-rank $p < 0.0001$), where Cluster 1 again exhibited the worst RFS. A significant difference in RFS was also observed among clusters in the Stage Ⅱ-Ⅲ patient subgroup (Fig. 5D, $p = 0.021$). However, the risk ranking shifted within this subgroup, with Cluster 4 showing the poorest prognosis.

Table 2
Univariable and multivariable Cox regression analysis of factors influencing RFS in surgical patients with classical EGFR mutations.

Variables	Univariable analysis		Multivariable analysis	
	HR (95%CI)	P value	HR (95%CI)	P value
Gender (Male vs. Female)	0.75(0.4–1.43)	0.385		
Age	1(0.97–1.04)	0.797		
Stage (vs. Stage Ⅰ)				
Stage Ⅱ	3.09(1.29–7.42)	< 0.05	1.86(0.75–4.61)	0.182
Stage Ⅲ	2.51(0.76–8.27)	0.129	2.15(0.64–7.22)	0.216
Stage Ⅳ	6.43(1.52–27.23)	< 0.05	5.51(1.14–26.67)	< 0.05
Cluster (vs. Cluster 1)				
Cluster 2	0.11(0.05–0.27)	< 0.001	0.12(0.05–0.28)	< 0.001
Cluster 3	0.12(0.06–0.27)	< 0.001	0.13(0.06–0.30)	< 0.001
Cluster 4	0.4(0.16–1.01)	0.053	0.37(0.13–1.03)	0.058

Analyais of OS across different co-alteration clusters

In the entire cohort, Cluster 1 and Cluster 4 exhibited significantly shorter OS compared to those in Clusters 2 and 3 (Fig. 6A, p = 0.014). While univariable Cox regression confirmed significant OS differences among clusters, multivariable analysis adjusting for tumor stage and other confounders showed that cluster assignment was not an independent prognostic factor for OS (Fig. 6B and Table 3).

Subgroup analysis revealed no significant difference in OS was observed among clusters in stage Ⅱ patients (Fig. 6C, p = 0.073). In contrast, a significant difference in OS was noted among clusters in stage Ⅲ-Ⅳ patients (Fig. 6D, p = 0.04), with Cluster 1 and Cluster 4 exhibiting a trend towards poorer survival.

Table 3
Univariable and multivariable Cox regression analysis of factors influencing OS in surgical patients with classical EGFR mutations.

Gender (Male vs. Female)	Univariable analysis		Multivariable analysis	
	HR (95%CI)	P value	HR (95%CI)	P value
Age	2.28(0.86–6.02)	0.096		
Stage (vs. Stage Ⅰ)	0.96(0.91–1.01)	0.137		
Stage Ⅱ				
Stage Ⅲ	5.67(1.48–21.71)	< 0.05	4.5(0.97–20.88)	0.054
Stage Ⅳ	6.87(1.78–26.5)	< 0.01	5.28(1.13–24.6)	< 0.05
Cluster (vs. Cluster 1)	70.97(15.84-317.88)	< 0.001	50.69(7.44–345.2)	< 0.001
Cluster 2				
Cluster 3	0.22(0.05–0.9)	< 0.05	0.41(0.09–1.84)	0.246
Cluster 4	0.25(0.08–0.83)	< 0.05	0.59(0.14–2.48)	0.469
Gender (Male vs. Female)	1.09(0.27–4.51)	0.901	0.8(0.11–5.71)	0.823

Disccussion

Based on NGS data from 494 patients, we systematically characterized the co-alteration landscape across various driver gene mutations and their subtypes in NSCLC. Furthermore, we developed and validated a molecular stratification system based on co-alteration profiles that demonstrated significant prognostic discrimination in surgical patients with classic EGFR mutations. Our study yielded several key findings: First, distinct co-alteration patterns were observed not only between different driver genes but also among subtypes within the same driver gene. Second, cluster analysis based on co-alteration profiles effectively identified surgical patients with classical EGFR mutations who exhibited varying risks of recurrence, and this prognostic value remained independent of the traditional TNM stage.

Co-alterations represent a fundamental aspect of tumor genomic complexity and provide a critical biological basis for oncogenic drivers to exert their tumorigenic functions and drive therapy resistance. Our findings confirmed that TP53 was the most prevalent co-alteration in NSCLC, consistent with previous research, highlighting the crucial role of TP53 pathway inactivation as a near-universal event in tumorigenesis[13, 14]. Particularly noteworthy was the highly specific co-occurrence of STK11 alterations with KRAS mutations, aligning with previous reports that the KRAS/STK11 co-alterations defined a

molecular subtype characterized by a distinct tumor microenvironment (e.g., reduced T-cell infiltration) and specific clinical outcomes (e.g., primary resistance to immune checkpoint inhibitors)[15, 16]. Additionally, the high frequency of RTK pathway co-alterations in ROS1-fusion tumors suggested a potential mechanism whereby parallel activation of other receptor tyrosine kinases sustains oncogenic signaling, possibly informing strategies to overcome targeted therapy resistance.

Finer subtyping of driver mutations revealed previously underappreciated genomic heterogeneity. Among EGFR mutated patients, exon 20 insertion showed higher co-alteration frequencies in DNA damage repair and RTK pathways compared to classical mutations. This finding provides a potential molecular mechanism for this subtype's relative insensitivity to first- and second-generation EGFR-TKIs: aberrations in DNA repair pathways may enhance genomic instability and accelerate the evolution of resistant clones, while activation of multiple RTK pathways may provide alternative survival signals, bypassing EGFR inhibition[17–19]. Similarly, among KRAS mutations, non-G12C exhibited more complex and diverse co-alteration profiles than G12C variants, suggesting potentially more heterogeneous oncogenic mechanisms, which indicated that KRAS mutations should not be considered a homogeneous entity in therapeutic development. Moreover, the generally higher co-alteration burden in BRAF V600E mutants may underlie their more aggressive tumor behavior..

The finding with the greatest potential for clinical translation is our successful conversion of co-alteration profiles into an applicable prognostic tool. In the large subgroup of surgical patients with classic EGFR mutations, unsupervised cluster analysis identified four distinct subgroups based on co-alteration profiles: Cluster 1 was characterized by the presence of TP53 alteration accompanied by limited other co-alterations. Cluster 2 exhibited scattered co-alterations lacking a dominant pattern, representing a heterogeneous, low-burden mutational signature. Cluster 3 demonstrated minimal additional genomic alterations, potentially defined as a genomically quiet subtype. Cluster 4 displayed a multi-gene co-alteration landscape involving DNA damage repair, RTKs, and drug transporter-related genes, suggesting a genomically complex phenotype.

Survival analysis confirmed Cluster 1 as a high-risk subgroup with the worst RFS in both the entire cohort and early-stage (stage Ⅰ) patients, independent of conventional clinicopathological factors. This aligns with the established biology: TP53 inactivation would disrupt cell cycle checkpoints, impair DNA repair, and promote genomic instability[20, 21]. In the context of surgical patients with classical EGFR mutations, TP53 co-alteration collectively constitutes a more aggressive biological subtype. It not only promotes initial tumor development and progression but also likely contributes to earlier disease recurrence and poorer survival outcomes by enhancing tumor plasticity, accelerating clonal evolution, and mediating tolerance to cytotoxic drugs and targeted therapies. Multivariable analysis revealed that its high-risk status was independent of traditional factors like age, gender, and TNM stage. Thus, for surgical patients with classical EGFR mutations in stage Ⅰ, who are commonly considered to have a favorable prognosis, being classified into Cluster 1 by NGS profiling may warrant more intensive surveillance or consideration of adjuvant targeted therapy. Interestingly, the highest-risk profile shifted to Cluster 4 in stage Ⅱ-Ⅲ patients, indicating that the prognostic impact of co-alterations may interact with tumor burden, the biology of which merits further investigation.

Although clusters strongly predicted RFS in multivariable analysis, its value for OS was attenuated. This phenomenon might be attributed to several factors: Firstly, diverse subsequent systemic therapies after disease recurrence (including multiple lines of TKI therapy, chemotherapy, anti-angiogenic therapy, and even immunotherapy) can significantly influence ultimate survival time, thereby diluting the direct impact of the initial co-alteration pattern on OS. Secondly, the absolute number of OS events in this study was relatively limited, potentially restricting statistical power. However, the observed trend towards OS differences in the Stage Ⅱ-Ⅲ subgroup still suggests that co-alterations might continue to influence prognosis in advanced stages.

Our study has several limitations. Firstly, the single-center, retrospective nature may introduce potential selection bias, necessitating validation in multi-center, prospective, large-scale cohorts. Secondly, variations in gene detection methods, NGS panels (covering genes, sequencing depth), and bioinformatic analysis pipelines might affect co-alteration calling, especially for low-frequency mutations, potentially impacting the homogeneity and precision of the results. Finally, the study primarily focused on the prognostic value of co-alterations; the potential impact of different co-alteration clusters on response to targeted therapy or immunotherapy remained unexplored. Future research can validate the robustness of our stratification system using standardized detection protocols in multi-center prospective cohorts, investigate its predictive value for response to targeted and immunotherapies, and elucidate the underlying molecular mechanisms to facilitate the model's translation into a clinical decision-making tool.

In summary, our systematic genomic analysis of NSCLC delineated a comprehensive co-alteration landscape across eight driver mutations and their subtypes, elucidating the complex genomic architecture underlying tumor heterogeneity. On this foundation, we established and validated a co-alteration-based prognostic model that enables precise risk stratification in surgical patients with classical EGFR mutations, particularly in early-stage disease, independent of TNM stage. This model offers a robust molecular framework to inform individualized adjuvant therapy decisions and underscores the clinical utility of co-alteration profiling in precision oncology.

Methods

Patients’ selection and data collection

Our study retrospectively enrolled 494 patients diagnosed with NSCLC between 2016 and 2023 at the First Medical Center of the Chinese PLA General Hospital. This study was approved by the Institutional Review Board (IRB) of the Chinese PLA General Hospital (Approval No. S2024-377-01). Due to the retrospective nature of this study, the requirement for written informed consent was waived by the ethics committee. All procedures performed in studies were in accordance with the ethical standards of the Declaration of Helsinki. All procedures performed in studies were in accordance with the principles of the Medical Ethics Committee. The inclusion criteria were: (1) histopathologically confirmed NSCLC, and (2) identification of at least one driver gene mutation through next-generation sequencing (NGS). Data on basic demographics (age, gender), tumor characteristics [stage (based on the AJCC 9th edition), pathological type], and sample source were retrieved from the electronic medical records.

Driver gene mutations were defined as follows:

EGFR: classical mutations (exon 19 deletion, L858R), exon 20 insertions and other uncommon mutations (L861Q, G719X, S768I and T790M, et al).

KRAS: G12C mutations and other non-G12C mutations (e.g., G12V, G12D, G13D).

BRAF: V600E mutations and other non-V600E mutations (e.g., K601E, G469A, D594G).

ALK, ROS1, or RET rearrangements (any partner gene).

MET: exon 14 skipping mutations, gene amplification, and other kinase domain activating mutations (including but not limited to Y1230C, D1228N).

HER2: exon 20 insertions.

The detailed patient enrollment process and exclusion criteria are illustrated in Fig. 7.

Genetic testing and functional pathway categorization of co-alterations

All enrolled patients underwent next-generation sequencing (NGS) using formalin-fixed paraffin-embedded (FFPE) tumor specimens with pathologically confirmed tumor cell content ≥ 20%. The study employed multiple validated NGS panels, each covering a minimum of 50 genes implicated in lung carcinogenesis, targeted therapy response, and immunotherapy susceptibility, including key driver genes such as EGFR, KRAS, BRAF, ALK, ROS1, RET, MET, and ERBB2. All sequencing procedures were performed in laboratories accredited by the College of American Pathologists (CAP) and/or certified under the Clinical Laboratory Improvement Amendments (CLIA), with a mean sequencing depth > 500× to ensure detection sensitivity for low-frequency variants.

Following quality control, raw sequencing data were processed using the Burrows-Wheeler Aligner (BWA) for sequence alignment and the Genome Analysis Toolkit (GATK) for variant calling. A comprehensive bioinformatics pipeline was employed for subsequent variant annotation. Detected variants were classified according to their clinical significance based on authoritative guidelines and large-scale genomic databases.

Based on biological pathway functions and previous evidence[22], co-alterations were systematically categorized into nine major functional classes: TP53 alterations, STK11 alterations, MYC pathway alterations, PI3K pathway alterations, DNA damage repair pathway alterations, β-catenin pathway alterations, cell cycle pathway alterations, receptor tyrosine kinase (RTK) pathway alterations, and other alterations (Table 4).

Table 4
The functional pathway categorization of co-alter

Functional pathway*	Co-alterations
TP53	TP53
STK11	STK11
MYC	MYC,MYCL,MYCN,MAX
PI3K	PIK3CA,PIK3R1,PIK3CB,AKT1,AKT2,AKT3,MTOR,RICTOR,TSC1, TSC2,PTEN
DNA repair	BRCA1,BRCA2,PALB2,RAD50,RAD51B,ATM,ATR,ATRX,MLH1,MSH2,MSH3,MSH6,TET2,DNMT3A,PMS1,PMS2
β-catenin	CTNNB1
Cell cycle	CDKN2A,CDKN2B,CCND1,CCND3,CCNE1,CDK4,CDK2,CDK6,CDK12,RB1
RTKs	EGFR,KRAS,MET,ALK,BRAF,ROS1,RET,ERBB2,ERBB3,ERBB4,KIT,PDGFRA,PDGFRB,FGFR1,FGFR2,FGFR3,FGFR4,NTRK1,NTRK2,NTRK3,TRK3,MAF
Others	PITCH1,ARID1A,SETD2,RNF43,SMAD4,NOTCH1,NOTCH2,NOTCH3,NOTCH4,IDH1,IDH2,NF1,NF2,CREBBP,SMARCA4,U2AF1,HNF1A,HIST1H3B,E L1(CD274),CTLA4,B2M,JAK2,TGFBR2,CXCL8,IL6

*Co-alteration in a given functional pathway was defined as the presence of at least one co-alteration within the pathway.

Statistical analysis

Analysis of clinical characteristics

Clinicopathological data were extracted from the electronic medical records. Categorical variables were presented as frequencies (percentages), with inter-group comparisons performed using χ^2 or Fisher's exact tests (when expected frequency was < 5). The normality of continuous variables was assessed using the Shapiro-Wilk test. Normally distributed variables were described as mean ± standard deviation and compared using Student's t-test or analysis of variance (ANOVA). Non-normally distributed variables were described as median (interquartile range) and compared using the Mann-Whitney U test or Kruskal-Wallis H test.

Co-alteration frequency and unsupervised cluster analysis

Stacked bar plots were employed to visualize the composition of co-alterations across functional pathways. Based on the predefined nine functional pathway categories, we calculated the frequency of patients with co-alterations in each functional pathway for every driver gene mutation and its respective subtypes. The frequencies of all functional pathways within each driver gene category were then stacked to show the cumulative frequency. In the resulting stacked bar

plot, the X-axis denoted the driver gene categories, the Y-axis indicated the cumulative frequency, and distinct colors were used to differentiate the various functional pathway classes.

For surgical patients with classical EGFR mutations, a binary co-alteration matrix (0: wild-type; 1: mutated) was constructed using genes with mutation frequencies > 5%. Unsupervised hierarchical cluster analysis was performed to identify patient subgroups with distinct co-alteration patterns. The cluster was based on binary distance and applied the Ward.D2 linkage method to minimize within-cluster variance. The optimal number of clusters (ranging from 2 to 5) was determined by calculating the average silhouette score across each cluster solution, with the highest score indicating the best partition. The result of cluster analysis was visualized as a heatmap using the R package pheatmap, with co-altered genes colored in red and wild-type genes in white.

Survival analysis

Survival data were collected through electronic medical records and telephone follow-ups. The primary endpoints were RFS and OS. RFS was defined as the time from surgery to the first observed disease recurrence or distant metastasis. OS was defined as the time from surgery to death from any cause. Survival curves were generated using the Kaplan-Meier method and compared with log-rank tests. Multivariable Cox proportional hazards models, adjusted for age, gender and tumor stage, were used to assess the independent prognostic value of co-alteration profiles.

Statistical softwares

All statistical analyses were performed using SPSS (version 27.0) and R (version 4.5.1). A two-sided p-value < 0.05 was considered statistically significant.

Declarations

Code availability

The underlying code for this study is not publicly available but may be made available to qualified researchers on reasonable request from the corresponding author.

Competing Interests

All authors declare no financial or non-financial competing interests.

Funding

This work was supported by Beijing Natural Science Foundation (No.7222164).The funder played no role in study design, data collection, analysis and interpretation of data, or the writing of this manuscript.

Author Contribution

J. J. and Z. X. designed the study. J. J., J. W., and Yi L. completed the inclusion and exclusion of all patients. J. J., K. Z., Z. Z., X. T., and X. Q.collected and managed the data. W. P., H. H., and M. H. conducted survival follow-up. J. J., W. C., X. L., X. Z., and Yu L. performed the code implementation and statistical analyses. J. J. drafted the original manuscript. Z. X. provided critical intellectual input and supervised the entire study. All authors interpreted the data, critically revised the manuscript and approved the final version of the manuscript prior to submission.

Acknowledgments

Z.X is funded by Beijing Natural Science Foundation (No.7222164).

Data Availability

The datasets generated and analysed during the current study are not publicly available due to patient privacy and ethical restrictions but are available from the corresponding author on reasonable request.

References

1. TAN A C, TAN D S W. Targeted Therapies for Lung Cancer Patients With Oncogenic Driver Molecular Alterations [J]. J Clin Oncol, 2022, 40(6): 611–25.
2. LYNCH T J, BELL D W, SORDELLA R, et al. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib [J]. N Engl J Med, 2004, 350(21): 2129–39.
3. WU Y L, DZIADZIUSZKO R, AHN J S, et al. Alectinib in Resected ALK-Positive Non-Small-Cell Lung Cancer [J]. N Engl J Med, 2024, 390(14): 1265–76.
4. WOLF J, SETO T, HAN J Y, et al. Capmatinib in MET Exon 14-Mutated or MET-Amplified Non-Small-Cell Lung Cancer [J]. N Engl J Med, 2020, 383(10): 944–57.

5. XU Y, TONG X, YAN J, et al. Short-Term Responders of Non-Small Cell Lung Cancer Patients to EGFR Tyrosine Kinase Inhibitors Display High Prevalence of TP53 Mutations and Primary Resistance Mechanisms [J]. Transl Oncol, 2018, 11(6): 1364–9.
6. MCGRANAHAN N, SWANTON C. Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future [J]. Cell, 2017, 168(4): 613–28.
7. VOKES N I, CHAMBERS E, NGUYEN T, et al. Concurrent TP53 Mutations Facilitate Resistance Evolution in EGFR-Mutant Lung Adenocarcinoma [J]. J Thorac Oncol, 2022, 17(6): 779–92.
8. LEE J K, LEE J, KIM S, et al. Clonal History and Genetic Predictors of Transformation Into Small-Cell Carcinomas From Lung Adenocarcinomas [J]. J Clin Oncol, 2017, 35(26): 3065–74.
9. OFFIN M, CHAN J M, TENET M, et al. Concurrent RB1 and TP53 Alterations Define a Subset of EGFR-Mutant Lung Cancers at risk for Histologic Transformation and Inferior Clinical Outcomes [J]. J Thorac Oncol, 2019, 14(10): 1784–93.
10. JI H, RAMSEY M R, HAYES D N, et al. LKB1 modulates lung cancer differentiation and metastasis [J]. Nature, 2007, 448(7155): 807–10.
11. WEST H J, MCCLELAND M, CAPPUZZO F, et al. Clinical efficacy of atezolizumab plus bevacizumab and chemotherapy in KRAS-mutated non-small cell lung cancer with STK11, KEAP1, or TP53 comutations: subgroup results from the phase III IMpower150 trial [J]. J Immunother Cancer, 2022, 10(2).
12. ZEHIR A, BENAYED R, SHAH R H, et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients [J]. Nat Med, 2017, 23(6): 703–13.
13. LANE D P. Cancer. p53, guardian of the genome [J]. Nature, 1992, 358(6381): 15–6.
14. EISCHEN C M. Genome Stability Requires p53 [J]. Cold Spring Harb Perspect Med, 2016, 6(6).
15. LEE H M, MUHAMMAD N, LIEU E L, et al. Concurrent loss of LKB1 and KEAP1 enhances SHMT-mediated antioxidant defence in KRAS-mutant lung cancer [J]. Nat Metab, 2024, 6(7): 1310–28.
16. QIAN Y, GALAN-COBO A, GUIJARRO I, et al. MCT4-dependent lactate secretion suppresses antitumor immunity in LKB1-deficient lung adenocarcinoma [J]. Cancer Cell, 2023, 41(7): 1363-80.e7.
17. CHOUAID C, FILLERON T, DEBIEUVRE D, et al. A Real-World Study of Patients with Advanced Non-squamous Non-small Cell Lung Cancer with EGFR Exon 20 Insertion: Clinical Characteristics and Outcomes [J]. Target Oncol, 2021, 16(6): 801–11.
18. XU R, YU S, ZHU D, et al. hCINAP regulates the DNA-damage response and mediates the resistance of acute myelocytic leukemia cells to therapy [J]. Nat Commun, 2019, 10(1): 3812.
19. XIA Y, WANG K, ZHAO J, et al. Receptor Tyrosine Kinase Fusion-Mediated Resistance to EGFR TKI in EGFR-Mutant NSCLC: A Multi-Center Analysis and Literature Review [J]. J Thorac Oncol, 2025, 20(4): 465–74.
20. BASLAN T, MORRIS J P T, ZHAO Z, et al. Ordered and deterministic cancer genome evolution after p53 loss [J]. Nature, 2022, 608(7924): 795–802.
21. QI L, LI G, LI P, et al. Twenty years of Gendicine® rAd-p53 cancer gene therapy: The first-in-class human cancer gene therapy in the era of personalized oncology [J]. Genes Dis, 2024, 11(4): 101155.
22. PASSARO A, ATTILI I, RAPPA A, et al. Genomic Characterization of Concurrent Alterations in Non-Small Cell Lung Cancer (NSCLC) Harboring Actionable Mutations [J]. Cancers (Basel), 2021, 13(9).

Figures

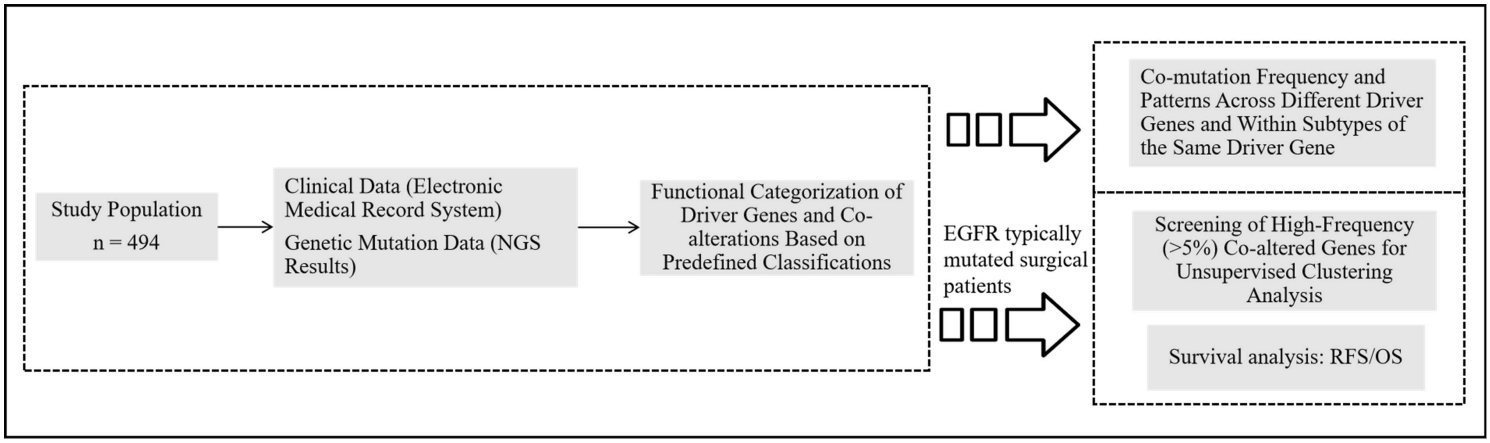
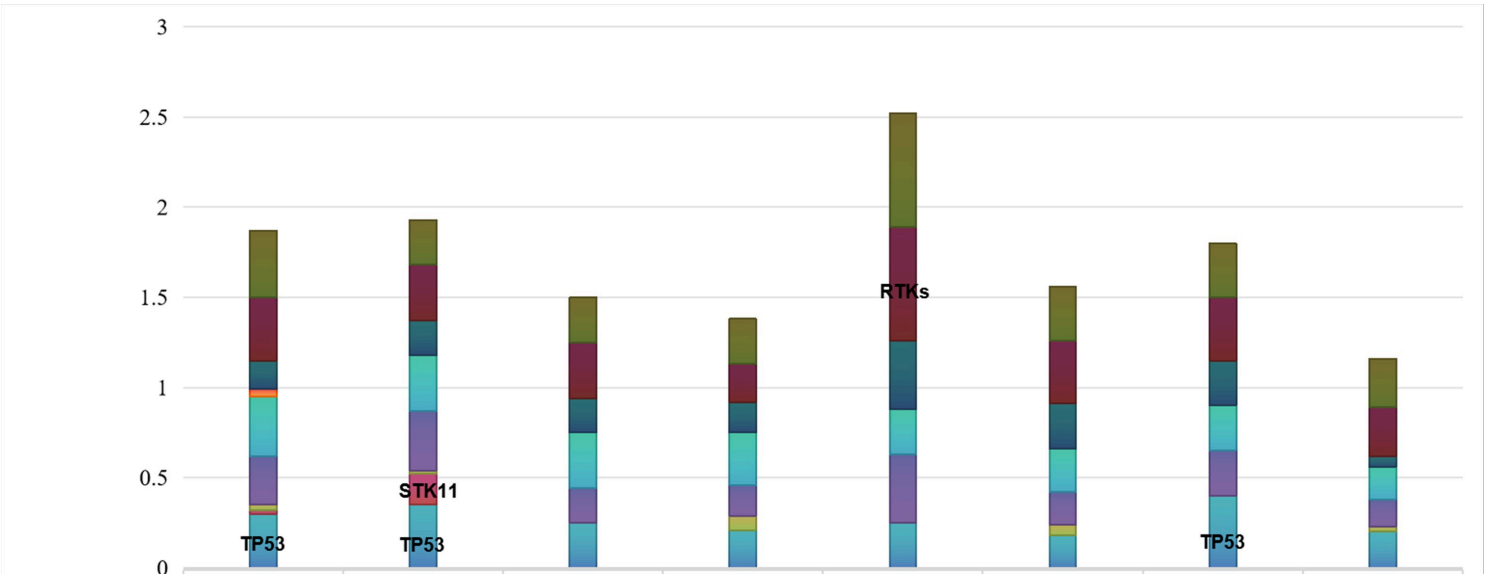


Figure 1

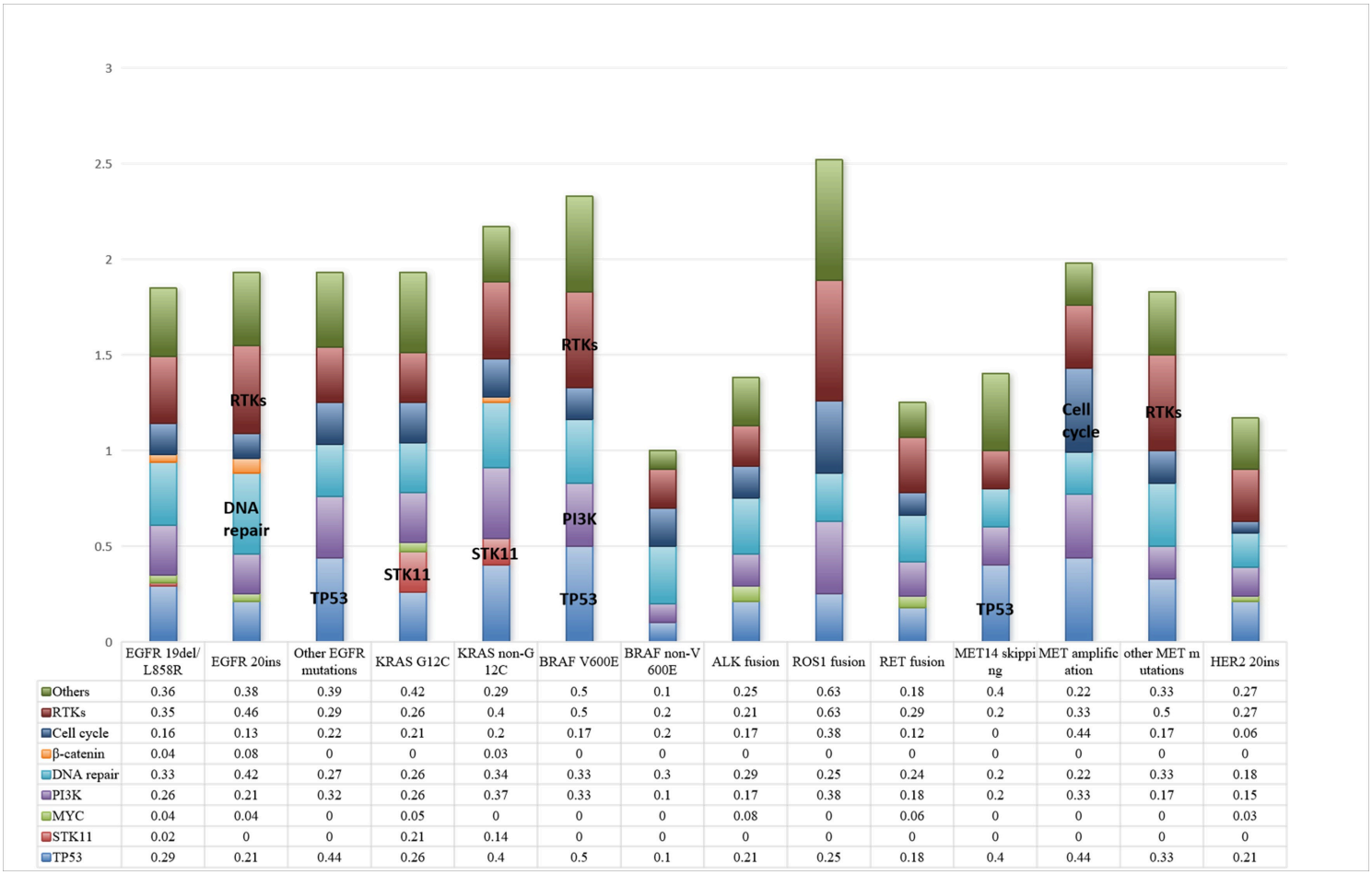
The flowchart of the study.



	EGFR	KRAS	BRAF	ALK	ROS1	RET	MET	HER2
Others	0.37	0.25	0.25	0.25	0.63	0.3	0.3	0.27
RTKs	0.35	0.31	0.31	0.21	0.63	0.35	0.35	0.27
Cell cycle	0.16	0.19	0.19	0.17	0.38	0.25	0.25	0.06
β-catenin	0.04	0	0	0	0	0	0	0
DNA repair	0.33	0.31	0.31	0.29	0.25	0.24	0.25	0.18
PI3K	0.27	0.33	0.19	0.17	0.38	0.18	0.25	0.15
MYC	0.03	0.02	0	0.08	0	0.06	0	0.03
STK11	0.02	0.17	0	0	0	0	0	0
TP53	0.3	0.35	0.25	0.21	0.25	0.18	0.4	0.2

Figure 2

Co-alteration frequencies across functional pathways in different driver mutations.



	EGFR 19del/ L858R	EGFR 20ins	Other EGFR mutations	KRAS G12C	KRAS non-G 12C	BRAF V600E	BRAF non-V 600E	ALK fusion	ROS1 fusion	RET fusion	MET14 skippi ng	MET amplific ation	other MET m utations	HER2 20ins
Others	0.36	0.38	0.39	0.42	0.29	0.5	0.1	0.25	0.63	0.18	0.4	0.22	0.33	0.27
RTKs	0.35	0.46	0.29	0.26	0.4	0.5	0.2	0.21	0.63	0.29	0.2	0.33	0.5	0.27
Cell cycle	0.16	0.13	0.22	0.21	0.2	0.17	0.2	0.17	0.38	0.12	0	0.44	0.17	0.06
β-catenin	0.04	0.08	0	0	0.03	0	0	0	0	0	0	0	0	0
DNA repair	0.33	0.42	0.27	0.26	0.34	0.33	0.3	0.29	0.25	0.24	0.2	0.22	0.33	0.18
PI3K	0.26	0.21	0.32	0.26	0.37	0.33	0.1	0.17	0.38	0.18	0.2	0.33	0.17	0.15
MYC	0.04	0.04	0	0.05	0	0	0	0.08	0	0.06	0	0	0	0.03
STK11	0.02	0	0	0.21	0.14	0	0	0	0	0	0	0	0	0
TP53	0.29	0.21	0.44	0.26	0.4	0.5	0.1	0.21	0.25	0.18	0.4	0.44	0.33	0.21

Figure 3

Co-alteration frequencies across functional pathways in different subtypes of driver mutations.

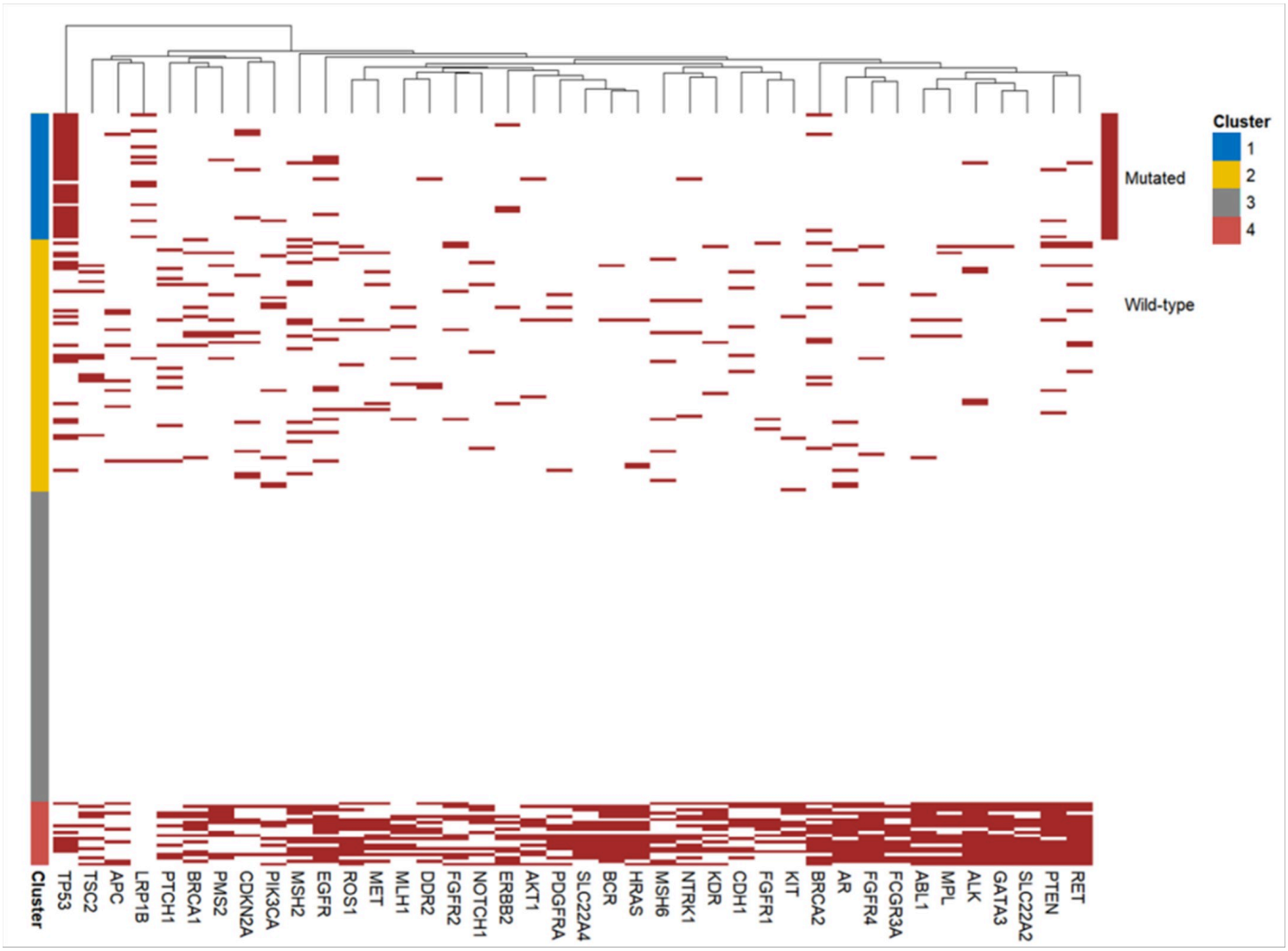


Figure 4

Cluster analysis on co-alteration profiles of surgical patients with classical EGFR mutations.

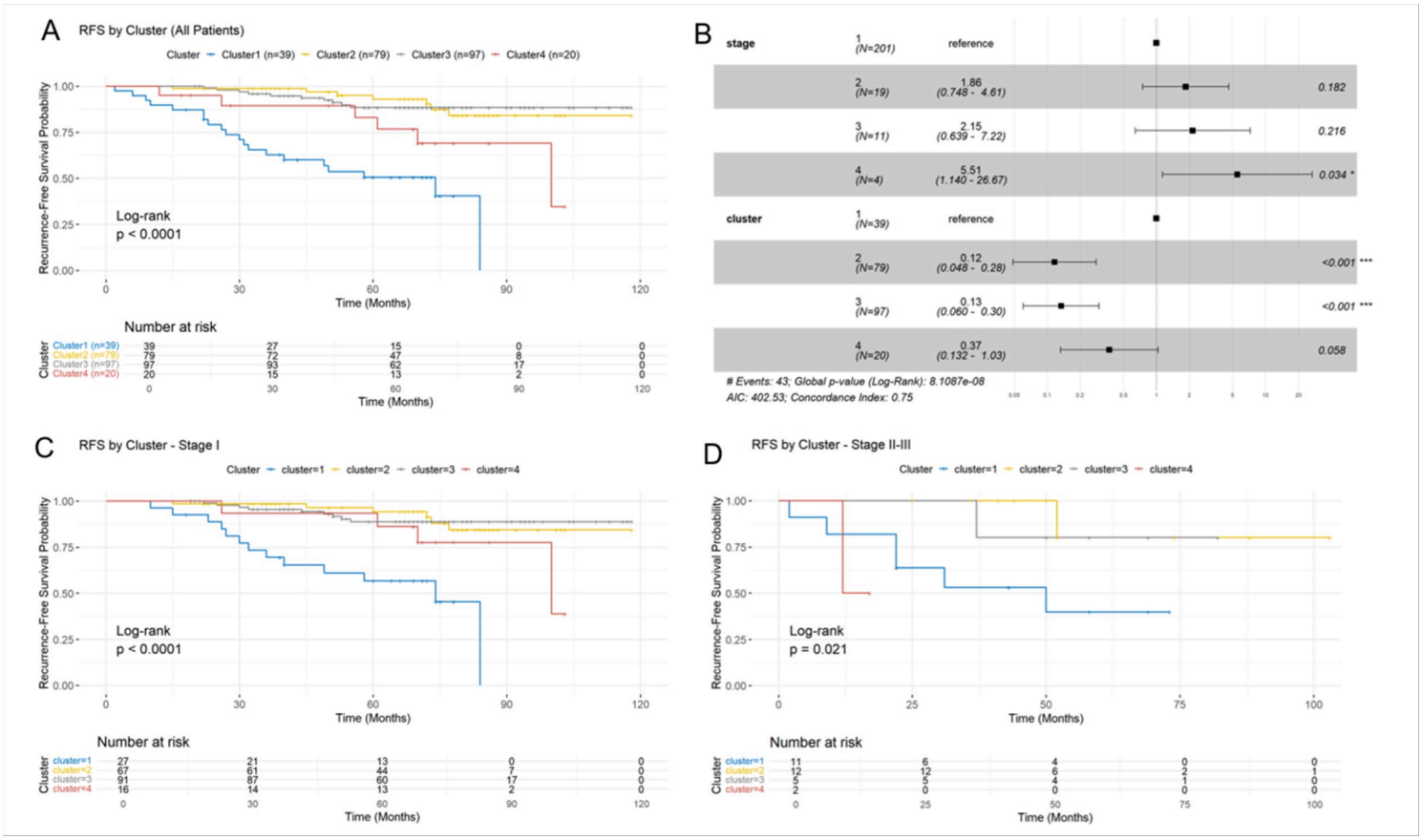


Figure 5

Prognostic value of co-alteration-based molecular subtyping for RFS in surgical patients with classical EGFR mutations. (A) Kaplan-Meier curves of RFS for distinct co-alteration-based clusters in the entire cohort. (B) Forest plot of the multivariable Cox regression analysis for RFS, incorporating tumor stage and co-alteration clusters. (C) RFS of different co-alteration clusters in the Stage Ⅰ patient subgroup. (D) RFS of different co-alteration clusters in the Stage Ⅱ-Ⅲpatient subgroup.

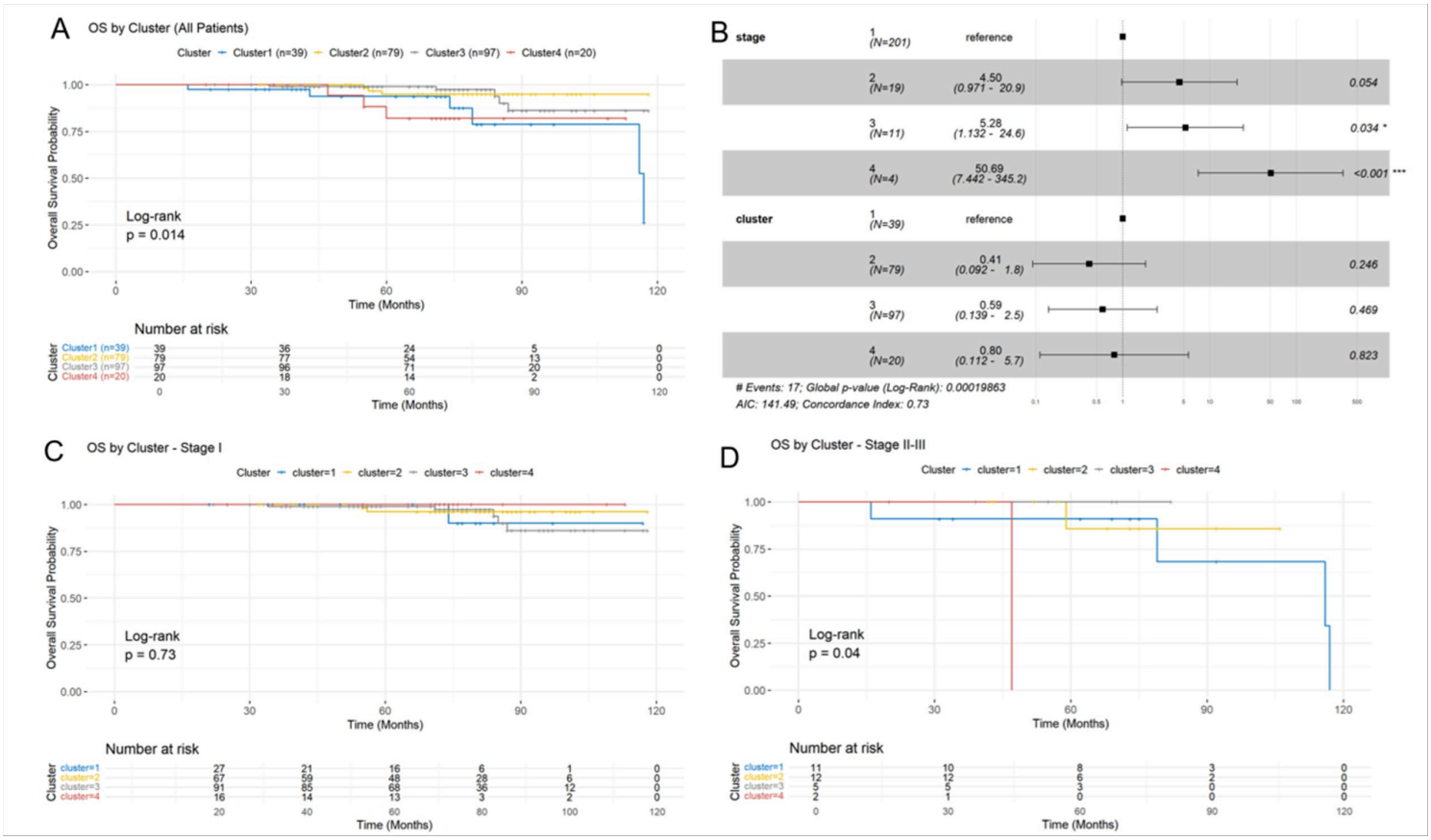


Figure 6

Prognostic value of co-alteration-based molecular subtyping for OS in surgical patients with classical EGFR mutations. (A) Kaplan-Meier curves of OS for distinct co-alteration-based clusters in the entire cohort. (B) Forest plot of the multivariable Cox regression analysis for OS, incorporating tumor stage and co-alteration clusters. (C) OS of different co-alteration clusters in the Stage Ⅰ patient subgroup. (D) OS of different co-alteration clusters in the Stage Ⅱ-Ⅲpatient subgroup.

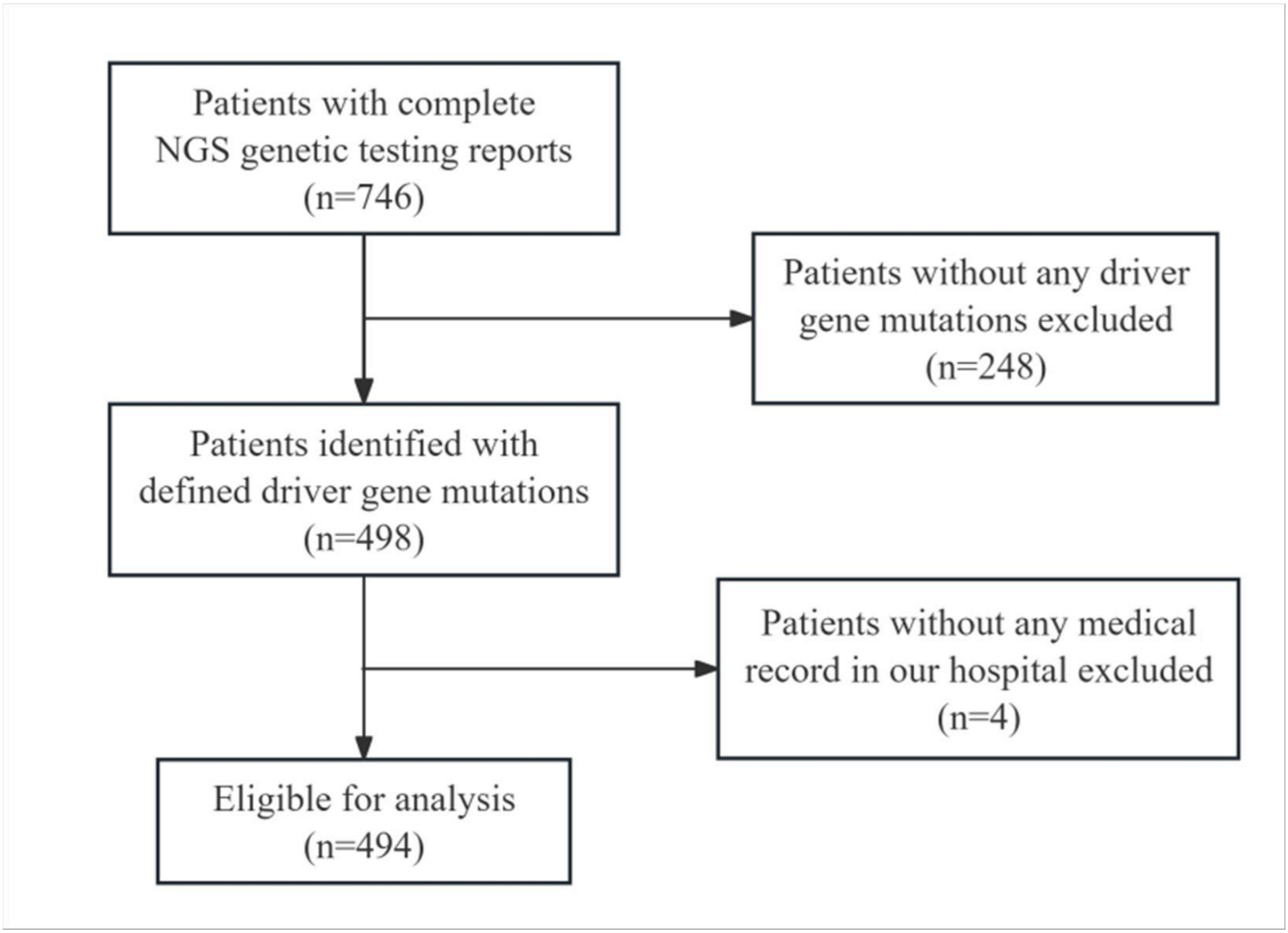


Figure 7

Flowchart of patients' selection. NGS, next-generation sequencing.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplement.docx